

スペクトルのローカルピークを用いた破裂子音と母音の認識に関する研究

著者	廖 莉莉
号	1238
発行年	1989
URL	http://hdl.handle.net/10097/6511

氏 名	廖 莉 莉
授 与 学 位	工 学 博 士
学位授与年月日	平成 2 年 3 月 28 日
学位授与の根拠法規	学位規則第 5 条第 1 項
研究科, 専攻の名称	東北大学大学院工学研究科 (博士課程) 情報工学専攻
学 位 論 文 題 目	スペクトルのローカルピークを用いた破裂 子音と母音の認識に関する研究
指 導 教 官	東北大学教授 城戸 健一
論 文 審 査 委 員	東北大学教授 城戸 健一 東北大学教授 木村 正行 東北大学教授 曾根 敏夫 東北大学助教授 牧野 正三

論 文 内 容 要 旨

第 1 章 序 論

従来の音声認識の研究では、振幅スペクトルに注目し、入力ボタンと標準ボタンの距離は振幅スペクトル間の距離で定義されることが多い。しかし、仮に振幅スペクトルが線スペクトルならば、ほとんどの場合正しい距離は得られない。また、スペクトル全体の情報を用いる場合には、伝送系の変動に弱いという問題がある。

一方、人間の場合、聴覚神経系における側抑制機能の存在や、LPF や HPF を通った音声でも十分に聞き取りが可能なおことから、人間が音声を聞いているとき、音声の振幅スペクトルではなく、振幅スペクトルのピークに対応する周波数に注目して聞き取りを行っていると考えられる。ホルマント周波数のような振幅スペクトルのピークに注目したパラメータの方が伝送系の変動に対応でき、個人性や調音結合による変動にも強く、正規化もしやすい。

従来、以上のような要請を満たす特徴量としてホルマントが良く用いられてきた。しかし、ホルマントは生成理論に基づくパラメータであり、正確なホルマントの抽出が困難という問題がある。また、聴覚系では、ホルマントより、スペクトルのピークが高次ニューロンに送られていると考えられる。そこで、我々は、聴覚側に立ったパラメータとして、スペクトルのローカルピークを特徴量として用いてきた。

ローカルピークを用いた音声認識では、まず、松岡、三輪、川端らの研究結果から、スペクトル

のローカルピークの周波数及びピークの周波数の動きを用いれば、高率で音声認識が行えることが明らかになっている。また金らは、スペクトルのローカルピークが、伝送系の変動に対処できることを、母音により実証した。そして、金らや古賀らの研究でも、スペクトルのローカルピークが、母音や子音の認識に有効なパラメータであることが示されている。

しかし、金らの研究には、まだ破裂時点の自動検出や認識率の向上などの幾つかの問題が残っており、さらに改善する必要がある。それらの問題点を解決するために、本研究では、①連続発声からの破裂時点の自動検出法の確立、②認識率を低下させずに、少ない計算量で破裂子音を認識するアルゴリズムの開発、③ローカルピークの抽出法の改善、を目的とした。

第2章 スペクトルの時間差分による破裂時点の自動検出

音声自動認識システムの実用化を考慮し、従来検討されていなかった破裂時点の自動検出について、チャンネルごとのスペクトルの時間差分値によって破裂時点を自動検出する方法を提案し、その有効性について検討した。

男女性各10名が発声した212単語中から切り出した無声破裂子音と有声破裂子音について、破裂時点の自動検出実験を行った。本章で提案したスペクトル差分法は従来のパワー差分法より優れている。ほぼ2,500Hz以上(21ch~29ch)の周波数帯域の各チャンネルの差分値の時間軸上のピークが、もっとも多く揃うフレームを破裂時点とすることによって、高率で破裂時点を自動検出できることが明らかになった。視察破裂時点の前後10フレーム、約210msの区間で、視察破裂時点に±2フレームのずれを許容した場合、無声破裂子音では96%、有声破裂子音では94%の精度で、破裂時点を検出できた。すなわち、約210msの長さの連続音声中の破裂時点の検出を、比較的簡単な方法で高精度に行い得るようになった。

第3章 直接法により抽出したローカルピークを用いた破裂子音の確認

直接BPF出力スペクトルから抽出したローカルピークを用いて、無声破裂子音と有声破裂子音の認識実験を行った。その際の尤度の計算法としては、1次のピークの出現確率とピークの出現しない確率を用いた。従来の同時確率より計算量がかなり減少する(約1/150)にもかかわらず、ほぼ同程度の認識率になることが分かった。その方法も用いて、不特定話者の男女声別と男女声混合について認識実験を行った結果、無声破裂子音では75%~79%、有声破裂子音では63%~70%の認識率が得られたが、まだ改善の余地がある。

認識結果を改善するために、ローカルピークを抽出するとき捨てていたスペクトルの情報を取り入れることを考えた。13ch(1000Hz)を分割チャンネルとして、スペクトルの低域と中高域の傾きを補助情報として、スペクトルのローカルピークと併用して認識実験を行った。その結果、無声破裂子音では86%~89%、有声破裂子音では74%~76%の認識率が得られ、スペクトルのローカルピークのみを用いる場合より約6%~16%の改善ができた。また、一般に伝送特性の変動に弱いと言われているスペクトルの傾きを、傾きの相対値を用いることによって、伝送特性の変化の影響を受けずに利用できることが確かめられた。そして、2章で提案した自動検出法により検出した破裂時点

を用いて認識実験を行った結果、視察破裂時点を用いたときの結果との差が小さいことが分かった。

以上の結果から、スペクトルの傾きが良いパラメータであることと、提案したスペクトルの時間差分による破裂時点の自動検出法は、破裂子音の認識のために有用であることが認識実験により明らかになった。

第4章 選択線形予測法によるローカルピークの抽出

直接法によるローカルピークの抽出法の問題としては、ピッチ周波数の高調波や、その他のホルマントに関係しないローカルピークが存在する場合が多かった。また、抽出したピークの中で、どれがホルマントに対応しているかを調べるのは難しいため、ローカルピークとホルマントの関係を示すのも困難であった。

それらの問題点を解決するために、本章では、スペクトルが3個の共振周波数を持つ全極モデルで表されるものと仮定して、 $Q \approx 6$ の帯域フィルタによる分析結果から、選択線形予測法によりローカルピークを抽出する方法を提案した。

その方法により抽出したローカルピークの出現頻度の周波数分布は、直接BPFの出力スペクトルから抽出したローカルピークより、余計なピークの少ないきれいな分布が得られた。選択線形予測法により抽出したローカルピークの母音五角形は、ホルマントの母音五角形より大きい五角形となっている。特に、 $/i/$ 、 $/e/$ が著しい。この性質は、母音を識別するには有利と考えられる。

第5章 選択線形予測法により抽出したローカルピークを用いた母音の認識

選択線形予測法によるローカルピークの抽出法の有効性を母音について検討した。従来ローカルピークを用いる場合、ピークの一次の出現頻度や同時確率を利用してきたが、ピークの付加や脱落が多かったため、ピークの相関やバンド幅などの利用は不可能だった。それに対し、4章で提案したローカルピークの抽出法は、スペクトルが3個の共振周波数を持つ全極モデルで表されるものと仮定しているので、抽出したピークの3個以下であり、ピーク間の相関や帯域幅などの利用も可能になる。

抽出したローカルピークの周波数及び帯域幅の情報の利用法について検討した結果、ピークの周波数及び帯域幅の逆数を用いたとき最も良い認識結果が得られた。男声では89%、女声では87%の認識率で、従来の直接法より約5%~9%の改善が得られたことから、選択線形予測法によるローカルピークの抽出法は、良い抽出法であることが示された。また、スペクトルの低・中高域の傾きの情報を併用することにより、さらに認識率が上がることも確かめられた。

第6章 選択線形予測法により抽出したローカルピークを用いた破裂子音の認識

選択線形予測法によるローカルピークの抽出法を破裂子音に対しても適用できるかどうかについて検討した。抽出したローカルピークの情報のみを用いて破裂子音の認識実験を行った結果、無声破裂子音では88%~89%、有声破裂子音では73%~76%の認識率が得られた。従来の直接法よりは、無声破裂子音では10%~13%、有声破裂子音では3%~13%の改善が得られた。また、スペクトル

の傾きと併用した場合、男女声別と男女声混合のいずれでも、従来の直接法より良い結果が得られている。無声破裂子音では89%~91%, 有声破裂子音では75%~78%の認識率になった。スペクトルの傾きを使わなくてもかなり高い認識率が得られたことから、破裂子音についても選択線形予測法により抽出したローカルピークは、有効なパラメータであると言える。

第7章 結 論

各章の結果をまとめると、

1. チャンネルごとのスペクトルの時間差分の最大値を用いれば、高率で破裂時点を自動検出できる。
2. スペクトルの最小二乗近似直線の上にあるローカルピークと低・中高域の傾きの補助情報を用いれば、/c, p, t, k/では86%, /b, d, g/では74%以上の認識結果が得られた。また、自動検出破裂時点をを用いても、視察破裂時点使用時の認識率との差は小さい。
3. 認識に有用なピークだけを正しく抽出することを目指して、選択線形予測法によるローカルピークの抽出法を提案した。この方法により抽出したローカルピークの母音五角形は、ホルマントの母音五角形より、大きい五角形となっている。
4. ピークの相関と帯域幅の利用が可能になった。ピークのチャンネル番号と帯域幅の逆数を用いた母音認識では、男声89%, 女声87%の認識結果が得られた。
5. 母音だけではなく、破裂子音についても選択線形予測法により抽出したローカルピークは有効なパラメータであることが分った。

以上全体をまとめると、スペクトルの時間軸上の時間差分のローカルピークを使うことにより、破裂時点の正確な検出を可能とすると共に、スペクトルのローカルピークを抽出する新しい有効な方法を開発した。そして、それを用いて抽出したローカルピークにより、不特定の話者が発声する母音と破裂子音の認識が高率で行えることを示した。

審 査 結 果 の 要 旨

音声認識の研究は活発に行われているが、人間の聴取能力に匹敵するようなシステムの実現には、まだ多くの問題が残されている。その解決には、人間が用いている聴覚パラメータの利用が有用と考えられている。この観点から、著者は、聴覚的特徴パラメータであるスペクトルのローカルピークによる破裂子音と母音の認識について研究し、種々の新しい認識方法を提案し、その有効性の検討を行った。本論文はその成果をまとめたもので、全編7章よりなる。

第1章は序論である。

第2章では、チャンネルごとのスペクトルの時間差分の最大値を用いることによって、破裂時点を自動検出する方法を提案している。その結果、無声破裂子音と有声破裂子音の破裂時点検出で、従来にない非常に高い検出率を得ている。これは実用上有用な結果である。

第3章では、スペクトルのローカルピークのほかに、スペクトルの低・中高域の傾きを併用することによって、認識率が大きく改善されることから、スペクトルの傾きが有用なパラメータであることを明らかにしている。これは重要な知見である。

第4章では、認識に有用なピークだけを正しく抽出することを目的として、選択線形予測法によるローカルピークの抽出法を提案している。この方法により抽出したローカルピークの分布とホルマントの分布を比較することによって、ローカルピークが、母音認識に有用な性質を持っていることを述べている。これは、興味深い結果である。

第5章では、選択線形予測法によるローカルピーク抽出法の、母音認識への有効性を検討している。抽出したローカルピークのチャンネル番号と帯域幅の逆数を用いることによって、従来の方法より認識率が改善されることを示している。

第6章では、選択線形予測法によるローカルピーク抽出法と、スペクトルの低・中高域の傾きを用いて、破裂子音の認識を行い、非常に高い認識率を得ている。

第7章は結論である。

以上要するに本論文は、聴覚的特徴パラメータであるスペクトルのローカルピークによる破裂子音と母音の認識について研究し、チャンネルごとのスペクトルの時間差分による破裂時点検出法や、選択線形予測法によるローカルピーク抽出法などを提案し、その有効性を明らかにしたもので、情報工学並びに音響工学に寄与するところが少なくない。

よって、本論文は工学博士の学位論文として合格と認める。